

# Classification Of Days Using Weather Impacted Traffic In The National Airspace System

Avijit Mukherjee<sup>1</sup>

*University Affiliated Research Center, Moffett Field, CA, 94035, USA*

Shon Grabbe<sup>2</sup> and Banavar Sridhar<sup>3</sup>

*NASA Ames Research Center, Moffett Field, CA, 94035, USA*

**Classification of days based on weather impact on the National Airspace System is essential to evaluate the effectiveness of traffic management decisions in the past, which ultimately can improve the operational readiness when similar events occur in the future. To achieve this goal, this paper presents a methodology to classify days based on severe weather impact on traffic. A daily index of the impact of severe weather on scheduled traffic flow, termed as the Weather Impacted Traffic Index, is used as an input to perform the classification. First, a factor analysis is performed to identify the dominant weather patterns that occur on various days. Six major weather factors are identified based on this analysis. Factor scores are obtained for each day based on the day's weather location and severity. Days are clustered using Ward's minimum-variance method applied to the daily factor scores. The outcome of the analysis is a set of 21 clusters and days within each cluster. While the weather and traffic in the days belonging to a common cluster are similar, they are not completely identical. Following the classification of days, the reroute advisories are analyzed to identify the frequently used routes on days belonging to various clusters. It is observed that the most frequently used reroutes on days that belong to a particular cluster exhibit similarity to the National Playbook routes designed to mitigate weather impact on those days, which is an intuitive result that is supported by data analysis.**

## I. INTRODUCTION

Adverse weather reduces the capacity of the National Airspace System (NAS) by partially or completely blocking routes, waypoints, and airports. During such conditions, traffic managers at the FAA's Air Traffic Control System Command Center (ATCSCC) and dispatchers at various Airlines' Operations Center (AOC) collaborate to mitigate the demand-capacity imbalance caused by weather. The end result is the implementation of a set of Traffic Flow Management (TFM) initiatives such as ground delay programs, reroute advisories, flow metering, and ground stops. The performance of the TFM control actions is measured using a set of metrics such as total delay, cancellations, diversions, additional flying time, airborne holding time, loss of predictability of operations, etc. These performance metrics vary from day-to-day based on the severity, location, and characteristics of weather as well as the effectiveness of TFM control actions.

If a particular day can be characterized as being similar, in terms of weather and traffic, to a few days in the past, then the TFM control actions from those days could serve as a basis for strategizing TFM on the current day of operation. A thorough post-operational evaluation of TFM actions in the past can reveal the potential areas of improvement, if possible. Doing so will better equip the NAS users (i.e., airlines) and the service provider (i.e., ATCSCC) with information to mitigate weather impact, and hence, improve the operational readiness. It will also improve the predictability of TFM control actions if the weather forecasts are reasonably accurate on a given day.

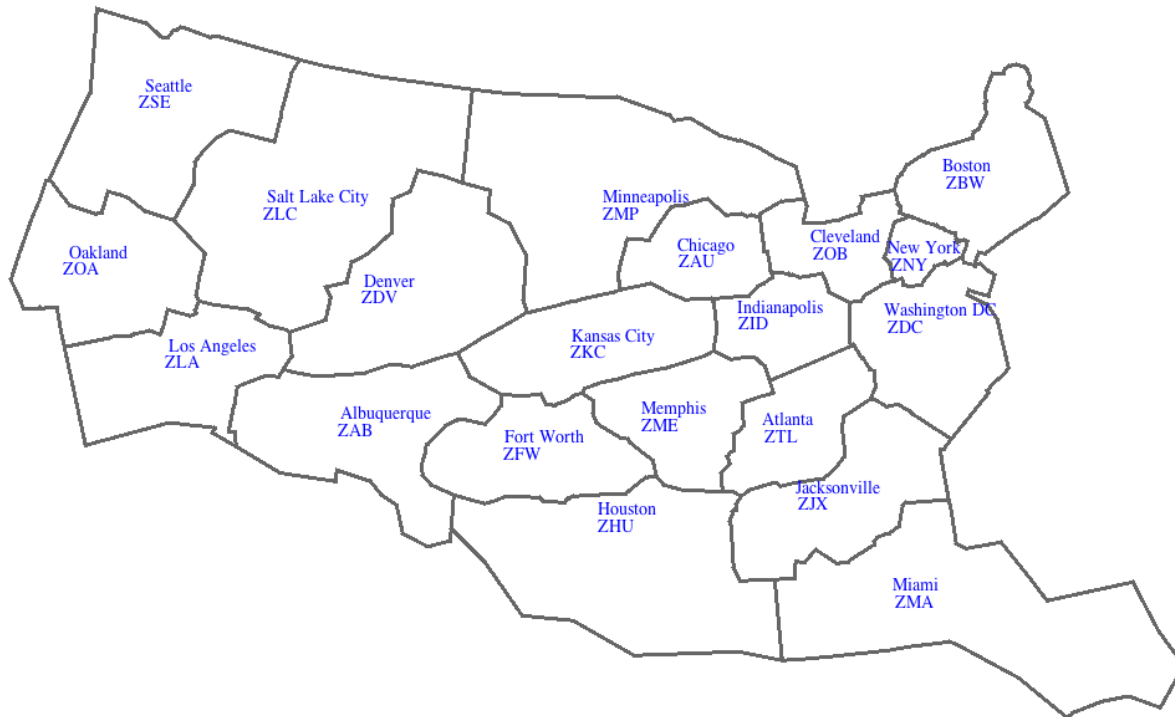
A successful classification of days is necessary to evaluate the effectiveness of TFM actions on days that are similar. This paper presents a methodology to classify days based on weather and traffic pattern and to cluster them into groups. Days belonging to the same cluster may not be identical, but are statistically close enough. The Weather Impacted Traffic Index (WITI) measures the location and severity of weather and its impact on traffic.<sup>1</sup> Daily WITI

<sup>1</sup> Research Scientist, MS 210-8.

<sup>2</sup> Research Scientist, Aviation Systems Division, MS 210-10, Associate Fellow of AIAA.

<sup>3</sup> Senior Scientist, Aviation Systems Division, MS 210-10, Fellow of AIAA.

of twenty Air Route Traffic Control Centers (commonly termed as Centers) encapsulating the continental United States (see FIGURE 1), and that of the entire NAS, for each day in 2011 are used to perform the classification. Using these data, a factor analysis<sup>2,3</sup> is performed to identify the dominant weather patterns. For each day, the dominant factors are scored based on the day's WITI values, which reveals the major weather phenomena on a given day. Clustering of days based on the factor scores is then performed.<sup>3</sup> The outcome of the analysis is a set of clusters and days within each cluster.



**FIGURE 1: Twenty Centers Encompassing Continental U.S. Airspace and their Abbreviations**

Center and NAS WITI used as input data primarily indicates the impact of severe weather on en route traffic, and to a lesser extent its impact on airports. Airport-level congestion measures<sup>4,5</sup> can be included to classify days even further, but this is left for future analysis. Following the classification of days based on WITI, the reroute advisories on days belonging to various clusters is analyzed. The purpose of this analysis is to identify if there are any commonly used reroutes on various cluster (i.e., type) of days, and whether those routes avoided the severe weather impacted regions.

The remainder of this paper is organized as follows. The next section presents the formal description of WITI, and a literature review of data classification methods applied to problems in air traffic management. Description of data is presented next. The following section presents the factor analysis and clustering results. The paper ends with conclusions followed by references.

## II. BACKGROUND

### A. Weather Impacted Traffic Index (WITI)

For the sake of readability, a brief description of WITI and its use in air traffic management research is presented here. WITI, as the name suggests, is an index that measures the impact of severe weather on traffic. On weather-impacted days aircraft are rerouted so they avoid severe weather. WITI for a weather-impacted day is computed by superimposing the weather cells over aircraft trajectories from a nominal (i.e. weather-free) day. Thus, WITI is an indicator of the severity of weather as well as its impact on nominal, weather-free, traffic. Suppose a grid is overlaid on the NAS, and  $W_{i,j}(t)$  be a binary variable, which assumes the value of 1 if the cell  $(i,j)$  in the grid is impacted by severe weather at time  $t$ ; 0 otherwise. Let  $T_{i,j}(t)$  denote the number of aircraft present in the cell  $(i,j)$  at time  $t$ .

The WITI of the cell  $(i, j)$  at time  $t$ , denoted as  $WITI_{i,j}(t)$ , is simply the product of  $W_{i,j}(t)$  and  $T_{i,j}(t)$ . WITI of a region (e.g., a Center) over a period of time is computed by summing the WITI of all grid cells that belong to the region over the time duration. A formal description of WITI is provided in Ref. 1. WITI is used in this paper primarily captures the impact of severe weather on en route traffic. This variant of WITI is used in the analyses presented in this paper. Other variants that compute airport-level impact of weather also exist in the literature.<sup>5</sup>

## B. Literature Review of Data Classification

There are well-established methodologies for data classification in general.<sup>6</sup> However, the literature on classifying days in the NAS based on weather impact is sparse. The relevant ones to this study are discussed in this section.

Ref. 7 applied principal component analysis to identify eight key variables that define the characteristics of a day. These are: total NAS delay, gate delays, on-time performance, traffic volume, cancellations, airport performance metric, volume-related delays, and weather-related departure delay minutes. They classified days based on these variables and identify six clusters. Days within a cluster have similar attributes as the key variables, and are representative of a *type* of day in the NAS. While the methodology for classifying days in Hoffman et al. and in this study are similar to a large extent, they differ in the selection of variables to characterize days. The key variables that characterize days in Ref. 7 are related to the performance metrics, which are the outcome of causalities and control actions taken. In the present study, however, the emphasis is to group days based on spatial distribution of weather impact. Thus, the variables used in this study are related to causality, rather than outcome, of disruption.

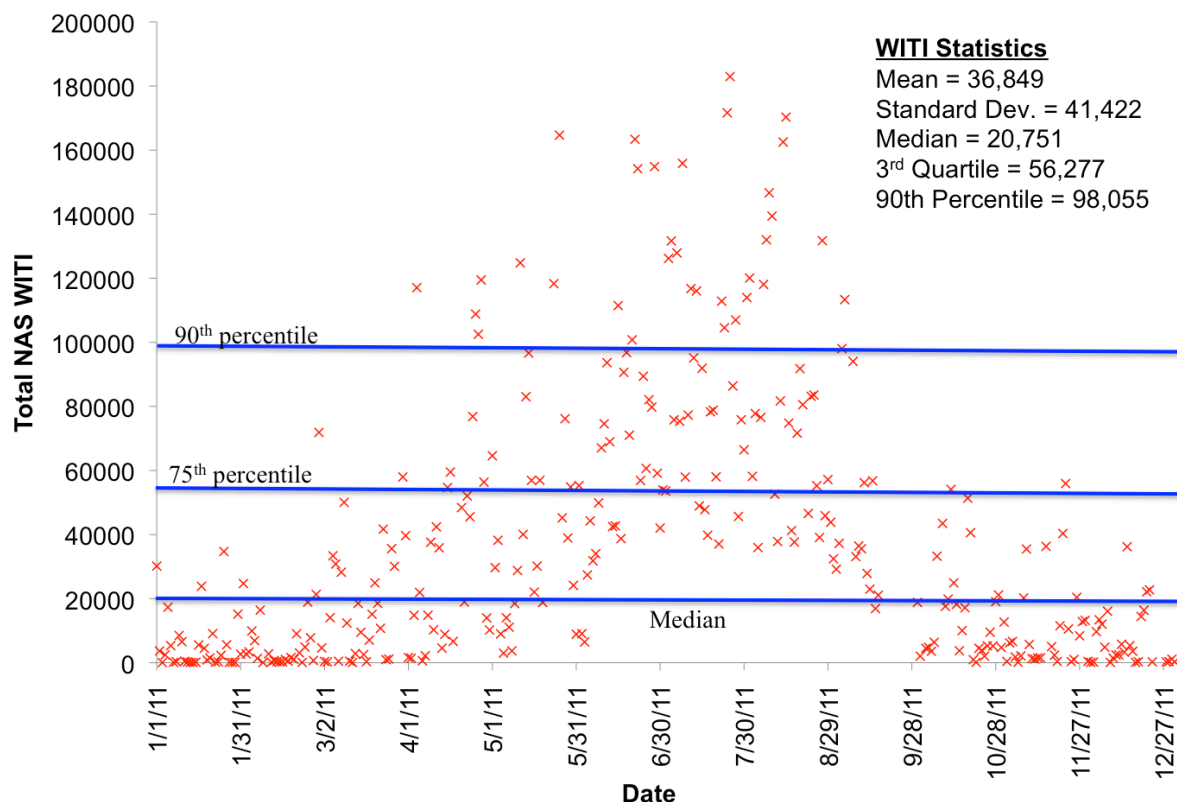
Another study closely related to this paper is by Ref. 8. This study divides the NAS into five regions and cluster days based on their WITI values. A K-Means clustering algorithm<sup>6</sup> is applied to cluster days. The primary difference between Ref. 8 and this study is the application of factor analysis that precedes clustering of days. Factors capture the correlation between WITI of various Centers, which is missing in Ref. 8. Moreover, using Center WITI in our study accounts for the spatial distribution of weather to a greater extent.

There are other applications of clustering and data classification in problems related to air traffic management. Ref. 9 applied a K-Means clustering algorithm to identify patterns of airport acceptance rates at a few of the busiest airports in the United States. Ref. 10 applied clustering methods to identify dominant flows at airport terminal areas as well as in the en route airspace, from a given set of individual flight trajectories.

## III. DATA DESCRIPTION

Daily WITI values for 20 Centers, shown in FIGURE 1, and for the entire NAS were obtained for 332 days in 2011. There were missing weather data for extended periods of time on 33 days, and hence, those days were omitted from the analysis. Traffic data for WITI computation was obtained from the Enhanced Traffic Management System (ETMS). As mentioned before, flight trajectories from weather-free days were used for WITI computation. The Corridor Integrated Weather System (CIWS), obtained from MIT Lincoln Lab, provided the location and severity of weather. The CIWS data provides Vertically Integrated Liquid (VIL) level at each cell in a 1-square kilometer grid overlaid on the continental U.S. VIL level greater than or equal to 3 was assumed to be severe. The Future Air Traffic Management Concept Evaluation Tool (FACET),<sup>11</sup> which is a simulation software developed at NASA, was used to compute WITI using the traffic and weather data. For the analyses in this paper, a day constituted of 24 hours ranging from 08:00 Coordinated Universal Time (UTC) to 07:59 UTC the following day.

FIGURE 2 shows the variation of daily total WITI for the entire NAS for the year 2011. A seasonal trend is evident from the figure. Convective weather during summer months increases WITI, whereas the weather impact is relatively low during winter, fall and spring months. Almost all days when the WITI is above the 3<sup>rd</sup> quartile are in the summer. As revealed by the descriptive statistics shown in FIGURE 2, there is noticeable day-to-day variation in WITI.



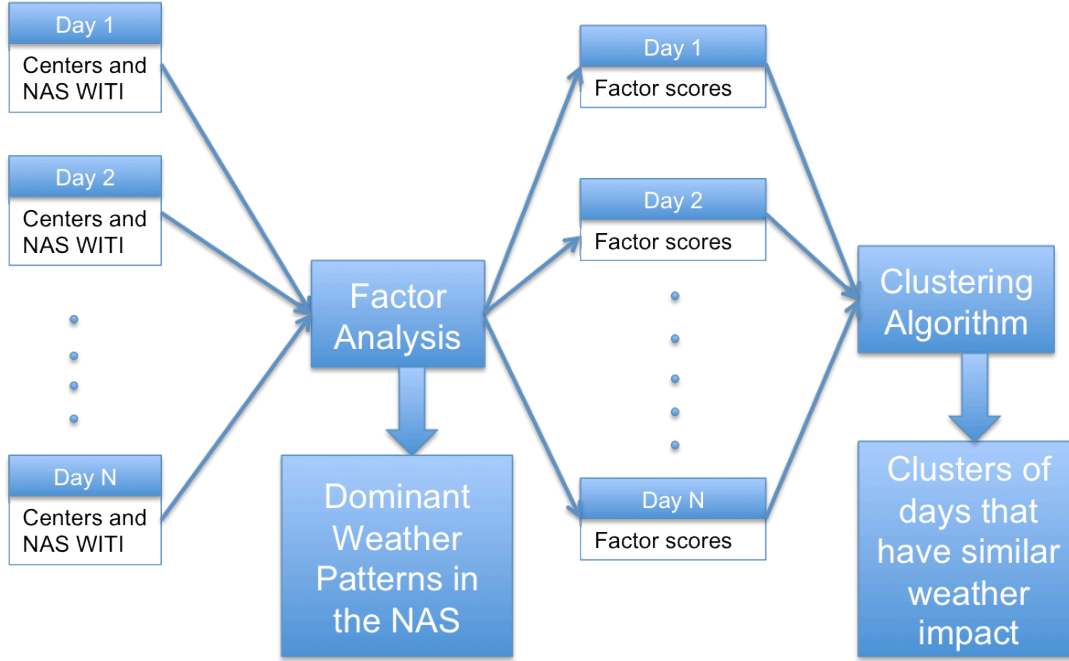
**FIGURE 2: Daily Total WITI of the NAS in 2011**

While the NAS WITI reveals the severity of weather impact over the entire continental U.S., the Center-level WITI serves as indicators of which Centers are impacted most. The five Centers that had highest average daily WITI are: Jacksonville (ZJX), Washington D.C. (ZDC), Atlanta (ZTL), Miami (ZMA), and Memphis (ZME). As shown in FIGURE 1 these Centers cover most of the northeast and southeast regions of the United States. Along with being impacted by severe weather, these Centers handle a high volume of traffic, which drives up the WITI.

Rerouting, along with airspace flow programs, are the main TFM controls implemented in response to en route weather. Most of the time, reroutes are selected from the National Playbook,<sup>12</sup> which provides a set of rerouting schemes for common adverse weather scenarios in the NAS. Reroute advisories for each day in 2011 were obtained from the FAA's National Traffic Management Log database. While the TFM controls were not used as a basis to classify days, the reroute advisories were analyzed to identify if there are any patterns of reroutes used on different types (i.e. clusters) of days.

#### IV. CLASSIFICATION OF DAYS USING WITI

As the total NAS WITI in a given day is the sum of that of the 20 Centers (shown in FIGURE 1), it suffices to include the WITI values of the NAS and any of 19 Centers in the factor and cluster analyses described here; Seattle Center (ZSE), which had the lowest average daily WITI, was excluded. FIGURE 3 describes the overall classification process. A factor analysis was performed using the daily WITI of the NAS and 19 Centers. The factor analysis generates the major weather factors, which are measured by location and severity of weather, impacting the NAS. For each day, factors are scored using the coefficients estimated from the factor analysis and the WITI values. The factor scores are then used as inputs to a clustering algorithm, which generates a set of clusters, each of which contains days that are similar in weather impacted traffic. SAS/STAT® Version 9.1 software was used to perform the analyses.



**FIGURE 3: Classification Process**

### A. Factor Analysis

Factor analysis is a method, which reduces the dimensionality of a dataset. This is a well-established method in data classification and there is substantial literature describing various aspects of factor analysis.<sup>2,3</sup> For the sake of readability, a brief description is provided here. Let  $\mathbf{X}^T = \{X_1, \dots, X_n\}$  be a vector of dimension  $n$ . In the present context  $\mathbf{X}$  represents a vector of dimension 20 (i.e., WITI of the NAS and 19 Centers). Let there be  $d$  observations of  $\mathbf{X}$ , which constitutes the input dataset to the factor analysis. Again, in the present context, each observation is the WITI values for the NAS and 19 Centers on a given day. The primary goal of the factor analysis is to relate the observed variables  $\mathbf{X}$  with a set of *latent* (i.e., unobserved) factors  $\mathbf{f}$ , called as *common factors*, as represented in Eq. (1). Let  $m$  be the dimension of  $\mathbf{f}$ . It is desired that the variability in the observed data can be accounted by a smaller set of factors (i.e.,  $m < n$ ).

$$\mathbf{X} = \mathbf{A}\mathbf{f} + \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (1)$$

Some of the important assumptions made during factor analysis are as follows. The common factors,  $\mathbf{f}$ , are assumed to be uncorrelated with each other and with the error  $\boldsymbol{\varepsilon}$ . The errors  $\boldsymbol{\varepsilon}$  are assumed to have mean 0, and are uncorrelated with one another. The vector  $\boldsymbol{\mu}$  represents the mean of  $\mathbf{X}$ . Based on Eq. (1), a component of  $\mathbf{X}$ , say  $X_i$ , can be represented in terms of the factors as in Eq. (2). The relationship between the variance of  $X_i$  (denoted by  $\sigma_i$ ), coefficients  $\lambda_{ij}$ , and the variance of the errors, which are denoted by  $\psi_i$ , is given by Eq. (3). The first component in the r.h.s. of Eq.(3),  $\sum_{j=1}^m \lambda_{ij}^2$ , is called the *communality* of variable  $X_i$ . It represents the extent to which the variance of  $X_i$  is accounted by common factors,  $\mathbf{f}$ . Larger values of this term indicates the factor model is able to account for the variability of the observed variables to a large extent.

$$X_i = \sum_{j=1}^m \lambda_{ij} f_j + \mu_i + \varepsilon_i \quad (2)$$

$$\sigma_i = \sum_{j=1}^m \lambda_{ij}^2 + \psi_i \quad (3)$$

There are several methods to estimate the coefficients  $\lambda_{ij}$ . The *principal factor* method has been applied in the current analyses, and is described here. The input to the method are the standardized observed variables and the corresponding correlation matrix  $\mathbf{R}$ . In the present context, each row of the input dataset contains WITI values of

the NAS and 19 Centers on a given day. Thus each column in the dataset is the WITI values of a given Center over multiple days. The mean and standard deviation of WITI of each Center (i.e., columns in the dataset) are used to standardize the variables. Using the standardized variables and the correlation matrix  $\mathbf{R}$ , Eq. (3) can be written in matrix form, as in Eq. (4). The goal is to estimate  $\mathbf{A}$  and  $\mathbf{\Psi}$ , given  $\mathbf{R}$ .

$$\mathbf{R} = \mathbf{A}\mathbf{A}^T + \mathbf{\Psi} \quad (4)$$

As a first step, the diagonal elements of the correlation matrix are replaced by the squared multiple correlation of  $X_i$  (standardized) with all other variables. This provides an estimate of  $\mathbf{R} - \mathbf{\Psi}$ , often called as the *reduced correlation matrix*.  $\mathbf{A}$  is then estimated by first decomposing the reduced correlation matrix using Spectral Decomposition Theorem,<sup>2</sup> and equating it to  $\mathbf{A}\mathbf{A}^T$ , as shown in Eq. (5). During this step, only  $m$  out of  $n$  eigenvalues, denoted by  $a_i$ s in the equation, of the reduced correlation matrix are used. Generally, the eigenvalues greater than 1 are chosen during this step. In the current analysis, six factors, whose respective eigenvalues were greater than 1, were retained. As indicated in TABLE 1, the six factors capture about 70% of the total variation in the input data.

$$\mathbf{R} - \mathbf{\Psi} = \sum_{i=1}^n a_i \mathbf{e}_i \mathbf{e}_i^T \cong \sum_{i=1}^m a_i \mathbf{e}_i \mathbf{e}_i^T = \mathbf{A}\mathbf{A}^T \quad (5)$$

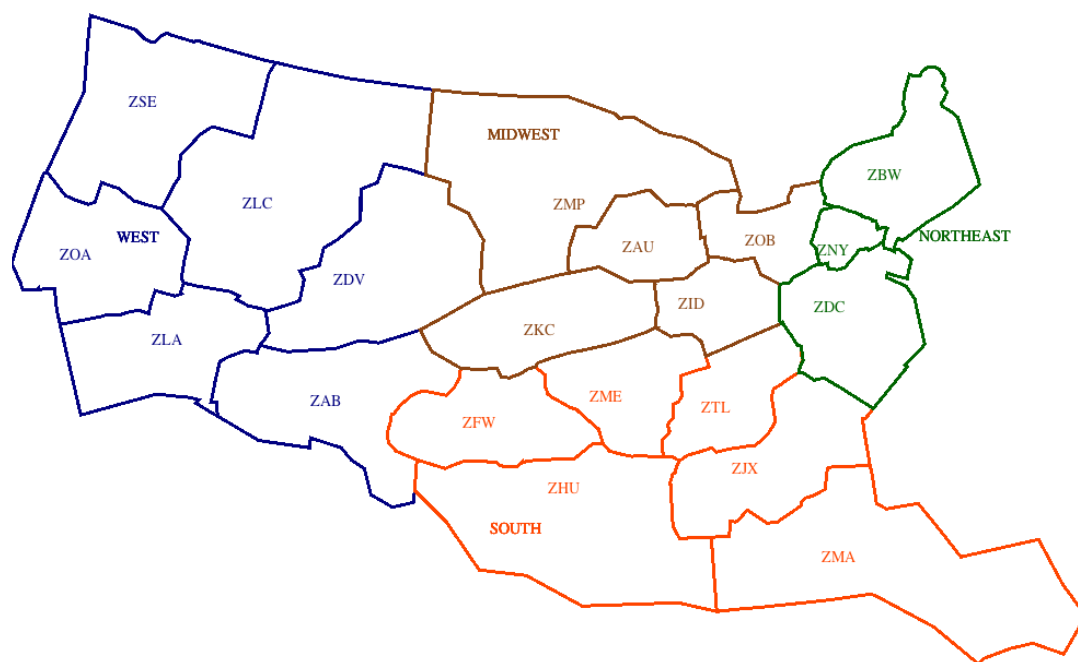
**TABLE 1: Description of Dominant Factors**

Factor	Eigenvalue	Cumulative Variation	Description
Factor 1	6.26	0.31	Severe weather impacting WITI of Centers in the southeast and western regions of the United States
Factor 2	2.45	0.44	Severe weather impacting the Northeast Centers
Factor 3	1.64	0.51	Severe weather impacting the Centers in the Midwest region
Factor 4	1.45	0.59	Severe weather impacting parts of Midwest and Atlanta Center (ZTL)
Factor 5	1.18	0.65	Mild weather in parts of the Western United States
Factor 6	1.05	0.70	Moderate weather impacting Centers in the south (ZFW, ZHU), and northern part of Midwest (ZMP)

If  $\mathbf{f}$  is set of factors that account for the variability of the observed variables  $\mathbf{X}$ , then any new set of factors, say  $\mathbf{f}^*$ , which are obtained by orthogonal transformation of  $\mathbf{f}$ , are also equally suitable to reduce the dimensionality of observed data. This can be easily verified by replacing  $\mathbf{f}$  by  $\mathbf{G}^T \mathbf{f}$  and  $\mathbf{A}$  by  $\mathbf{A}\mathbf{G}$  in Eq. (1), where  $\mathbf{G}$  is an orthogonal transformation matrix (i.e.,  $\mathbf{G}\mathbf{G}^T = \mathbf{I}$ ). In practice, the rotated factor loadings (i.e.,  $\mathbf{A}\mathbf{G}$ ) are used to better explain the relationship between the latent factors (rotated) and the observed variables. While there are several methods to achieve this, the *varimax* rotation method was chosen in the analyses of this paper. SAS/STAT® Version 9.1 software was used to perform the analysis. The rotated *factor loadings*, which indicate the correlation between observed variables (i.e., WITI) and rotated factors, are provided in TABLE 2. Based on the rotated factor loadings, the description of the six dominant weather factors are provided in TABLE 1. The numbers highlighted in bold in TABLE 2 indicate strong correlation (values higher than 0.40) between a factor and an input variable. FIGURE 4 shows four main regions in which various Centers belong, as described in the Census Bureau.<sup>13</sup> This geographic division of the Centers is used in describing various factors in the subsequent discussions.

**TABLE 2: Rotated Factor Loadings**

Regional WITI	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
NAS	<b>0.48</b>	<b>0.45</b>	<b>0.50</b>	<b>0.47</b>	0.14	0.26
ZAB	<b>0.44</b>	0.20	0.20	-0.01	<b>0.60</b>	0.15
ZAU	0.20	-0.07	<b>0.80</b>	-0.07	-0.11	0.12
ZBW	0.23	<b>0.82</b>	0.05	0.02	0.04	0.11
ZDC	0.17	<b>0.75</b>	0.07	0.20	0.28	-0.04
ZDV	<b>0.63</b>	0.18	0.16	-0.04	0.25	0.29
ZFW	-0.15	-0.05	0.18	0.24	0.05	<b>0.77</b>
ZHU	0.19	0.22	0.06	0.31	-0.11	<b>0.50</b>
ZID	0.02	0.11	<b>0.70</b>	<b>0.47</b>	0.03	0.07
ZJX	<b>0.79</b>	0.11	0.04	0.28	0.02	-0.06
ZKC	0.03	-0.02	<b>0.64</b>	0.11	0.24	0.30
ZLA	0.18	0.13	-0.03	-0.06	<b>0.81</b>	0.02
ZLC	<b>0.68</b>	0.23	-0.01	0.02	0.32	0.21
ZMA	<b>0.77</b>	0.13	0.16	0.09	-0.01	-0.12
ZME	-0.05	0.01	0.29	<b>0.80</b>	0.10	0.27
ZMP	<b>0.46</b>	0.01	0.25	-0.15	-0.04	<b>0.58</b>
ZNY	0.10	<b>0.94</b>	0.09	0.05	0.09	0.00
ZOA	-0.04	0.06	-0.01	0.19	<b>0.73</b>	-0.12
ZOB	0.12	0.34	<b>0.73</b>	0.18	-0.03	0.00
ZTL	0.37	0.20	0.06	<b>0.74</b>	0.07	0.06

**FIGURE 4: Geographic Division of the Centers into Four Regions**

The six dominant factors in TABLE 1 represent the most common weather impacted traffic patterns in the continental United States. WITI of the NAS and various centers are a linear combination of the factors. On a given

day, weather could span over multiple regions of the NAS triggering higher values for more than one factor. One of the important outcomes of the factor analysis is the estimate of *factor scores*. Higher score of a factor on a given day indicates the presence of the corresponding weather pattern on that day. Note that these are different from the factor loadings, which are the elements of  $\mathbf{A}$ . It is therefore of interest to estimate the factor scores, which represent the magnitude of various factors on each day, given the observed variables  $\mathbf{X}$  and the factor loadings  $\mathbf{A}$ . While there are several methods that estimate the factor scores, *Ordinary Least Squares* method,<sup>2</sup> which minimizes the sum of squares of the residuals, as given in Eq. (6), has been applied in the current analyses. Given the vector of observed variables on day  $i$ ,  $\mathbf{X}_i$ , the factor scores for that day,  $\mathbf{f}_i$ , are estimated by minimizing the squares of the residuals. It is these factor scores that are used as input to the clustering algorithm described in the next section.

$$\min_{\mathbf{f}_i} \mathbf{e}_i^T \mathbf{e}_i = \min_{\mathbf{f}_i} (\mathbf{X}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{f}_i)^T (\mathbf{X}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{f}_i) \quad (6)$$

## B. Clustering of Days

Ward's minimum-variance method<sup>14</sup> was applied to cluster days based on the factor scores on each day. Ward's algorithm is a hierarchical clustering method for reducing a set of observations into smaller subsets while minimizing the sum of variances of observations in each subset. In the present context, clusters are formed so that the total sum of squares of the Euclidean distances between factor scores of the days within the clusters is minimized. Twenty-one clusters were retained based on the pseudo- $t^2$  and pseudo-F statistics.  $R^2$  indicates that the 21 clusters account for almost 70% variation of the input data.

The severity of weather in various Centers and that in the NAS for the 21 clusters are shown in FIGURE 5. Severity of weather in the Centers is indicated by different colors. For a given cluster, the Centers whose daily WITI, averaged over all days in the cluster, is above the 90<sup>th</sup> percentile of the distribution of daily Center-level WITI are colored red. These are the Centers that face severe weather in the days that belong to the particular cluster. Similarly, yellow indicates WITI between 75<sup>th</sup> and 90<sup>th</sup> percentiles (i.e., moderate weather), while blue indicates mild or no weather.

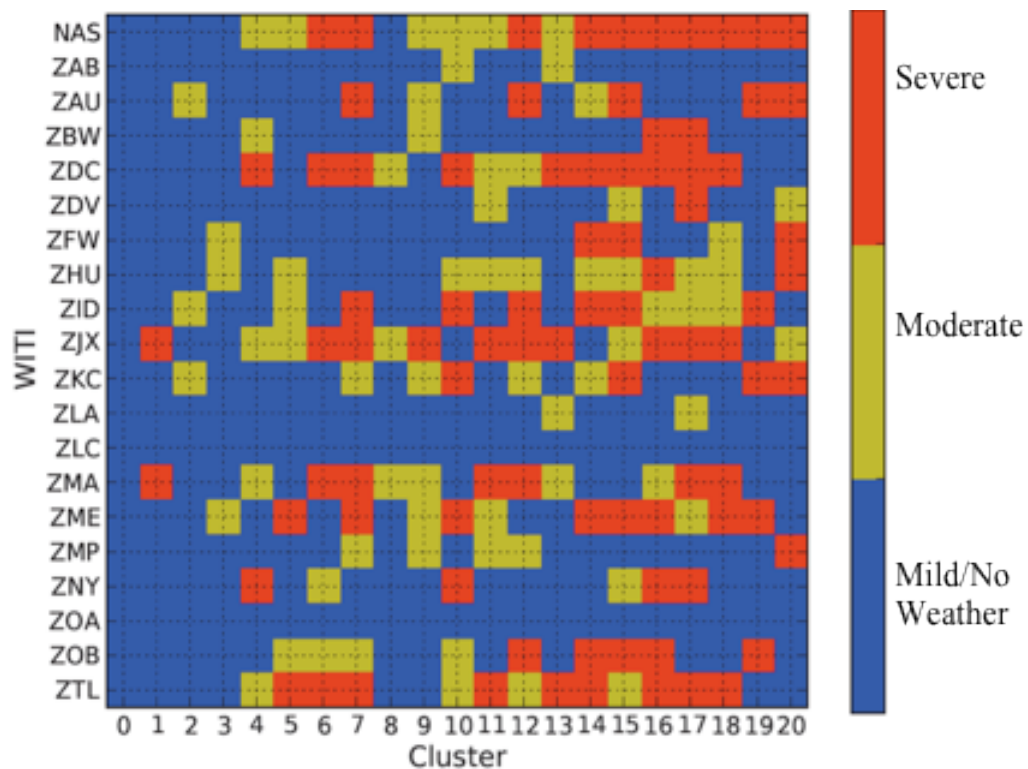
The overall NAS weather in the days belonging to the first four clusters (cluster numbers 0-3 in abscissa) is mild. However, there is variation in the Center-level WITI between the individual clusters. For example, Cluster1 indicates a mild weather day with primary impact on the southeast Centers – Jacksonville (ZJX) and Miami (ZMA). Cluster2, on the other hand, indicates days with weather in the Midwest impacting the following Centers: Chicago (ZAU), Indianapolis (ZID), and Kansas City (ZKC). Severity of overall NAS weather is high in the Clusters 6, 7, 12, and 14 – 20. Not surprisingly, there are many more Centers impacted by moderate and severe weather in these clusters compared to those corresponding to mild weather days. The number of days within each cluster varied between 1 and 123. There was only one day in each of the clusters 19 and 20. These days were unique to an extent that they did not belong to any other “groups” of days. Cluster0 contained 123 days, most of them in the winter, fall, and spring months. As shown in FIGURE 2, the NAS WITI in these months are usually low, which corroborates with FIGURE 5.

As indicated in FIGURE 5, there is noticeable variation in the Center-level WITI between the clusters. This is due to day-to-day variation in the weather, and in particular, the occurrence of multiple dominant weather factors described in TABLE 1. As discussed in the previous section, the dominant factors capture the correlation between Center-level WITI. As can be seen in FIGURE 5, the severity of weather in Jacksonville (ZJX) and Miami (ZMA) Centers is similar across clusters. This is because weather occurring in the southeast U.S. impacts both these Centers simultaneously.

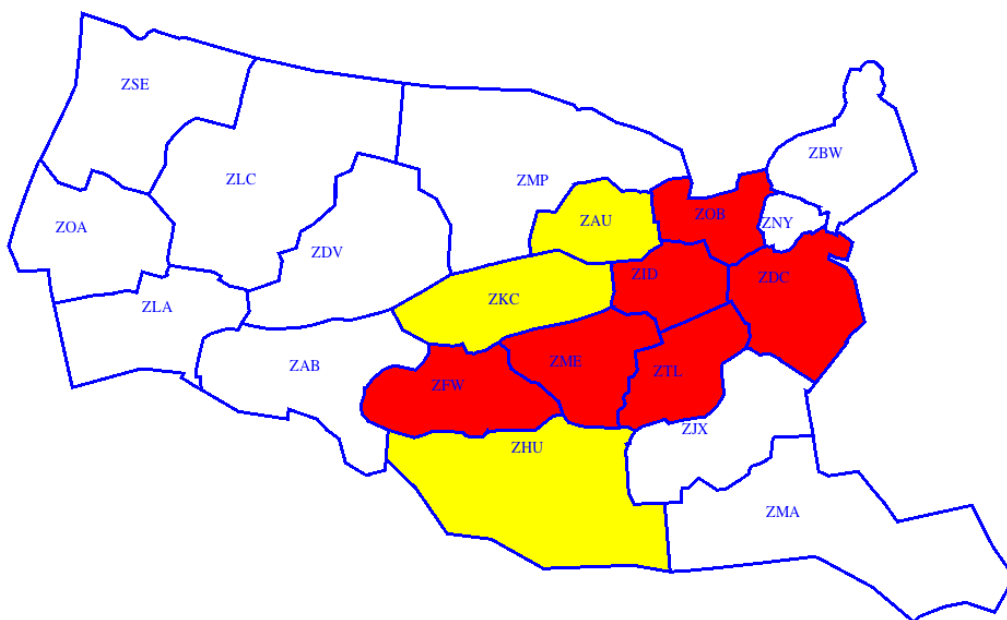
The Centers that are impacted by severe weather on the days belonging to Cluster14 are shown in FIGURE 6. There were 7 days in this cluster. As can be seen in the figure, severe weather primarily impacts Centers in the Midwest and southern United States. Weather factors 3, 4, and 6 had high scores on the days within this cluster. This corresponds to the Centers whose average WITI values are high in the figure. Reroute advisories on the days in this cluster were analyzed to identify the frequently used routes. FIGURE 7 shows 15 major National Playbook routes that were used to reroute aircraft in response to severe weather on the 7 days in this cluster. The top 15 Playbook reroutes constitute almost 50% of all the reroute advisories that were implemented on these days. The Playbook advisory “ATL\_NO\_BNA”, which constitute a set of reroutes for flights destined to Atlanta Hartsfield International Airport (ATL), was used most frequently. Reroutes for flights destined to Chicago O'Hare airport (ORD) were the second most commonly used. Several of the Playbook routes for flights using the Houston Center (ZHU) and those destined to Dallas Fort Worth International Airport (DFW) are used quite often. As evident from FIGURE 6, these airports and Centers are severely impacted on the days in Cluster14. Thus, not surprisingly, the frequently used



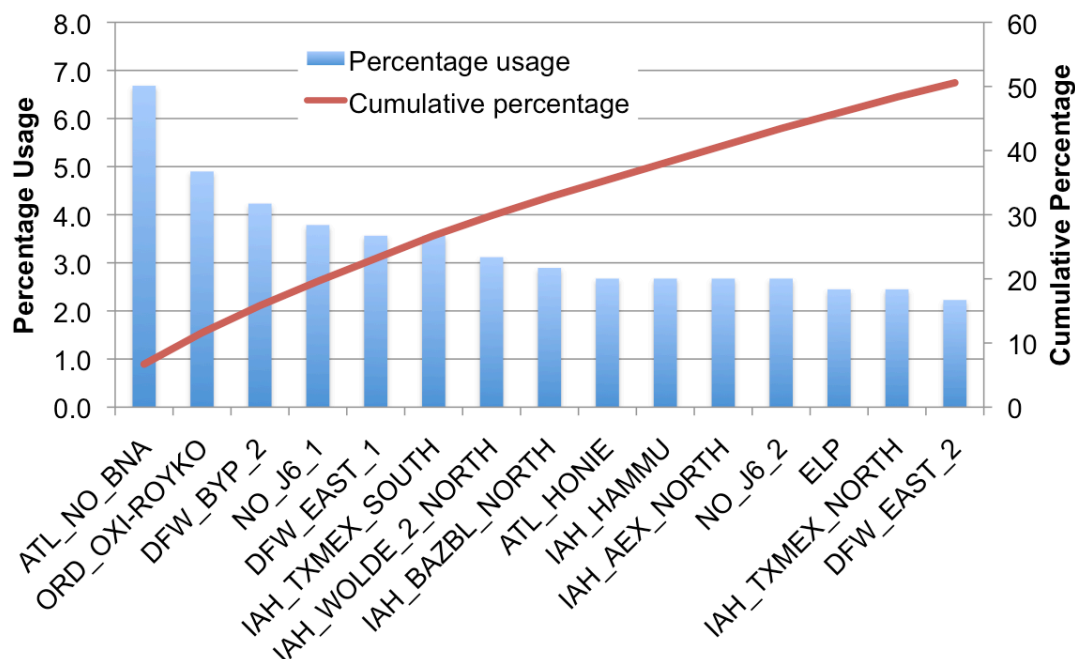
Playbook routes are the ones designed to avoid weather in the Centers that are severely impacted. Similar trends were observed for reroutes on days belonging to other clusters.



**FIGURE 5: Severity of NAS and Center-Level Weather for Various Clusters**



**FIGURE 6: Severe Weather Impacted Centers in Days Belonging to Cluster14**



**FIGURE 7: Top 15 National Playbook Reroutes Implemented in Days of Cluster14**

## V. CONCLUSIONS

This paper presents a methodology to classify days based on weather and its impact on traffic in the NAS. The daily WITI of twenty Centers encapsulating the continental United States, and that of the entire NAS, are used to perform the classification. Using these data, a factor analysis is performed that led to the identification of six dominant weather patterns in the NAS. For each day, the dominant factors are scored based on the day's WITI values and the scoring coefficients obtained from the factor analysis. The dominant factors represent the major weather phenomena that occur in the NAS on a typical day. Days are clustered using Ward's minimum-variance method applied to the daily factor scores. The outcome of the analysis is a set of 21 clusters and days within each cluster. Following the classification of days based on WITI, the reroute advisories on days belonging to various clusters are analyzed. It is observed that the most frequently used reroutes on days that belong to a particular cluster are the National Playbook routes designed to mitigate weather in the regions impacted severely on those days, which is an intuitive result that is supported by data analysis.

A distinguishing aspect of this research is the use of variables, which are the cause of disruption in the NAS, as input to the classification process. In the past, researchers have used performance metrics, which are the outcome of causal effect and controls, to do similar analysis. Classification of days into groups with similar causalities (i.e., weather and traffic) can facilitate researchers and TFM decision makers to analyze the effectiveness of TFM decisions that were made in the past. A thorough post-operational evaluation of TFM actions in the past can reveal the potential areas of improvement. It will prepare NAS users and decision makers to mitigate similar situations in the future, and hence, improve the operational readiness of the system. Another potential application of this research is to develop TFM strategies that are contingent upon realized weather scenarios. While the days belonging to a particular cluster are, in general, similar, they are not completely identical. Thus, by analyzing how weather evolved on different days of the same cluster, one can develop scenarios of weather, which can ultimately serve a basis for contingency planning of TFM decisions.

Future extension of this research will include airport-level congestion measures along with en route WITI to classify days. Larger numbers of days will be required to improve the performance measure of the factor analysis and clustering algorithm. Hourly WITI can be used to capture the varying impact of weather during different times of day. The paper is concluded with a note on the statistical methods used in the classification process. While there exists a plethora of methods for clustering data, the suitability of a particular method varies with the properties of the dataset on which it is applied.

## VI. REFERENCES

- <sup>1</sup>Sridhar, B. and N. Chen. Short-Term National Airspace System Delay Prediction using Weather Impacted Traffic Index. In *Journal of Guidance, Control, and Dynamics*, 32(2), 2009, pp. 675-661.
- <sup>2</sup>Mardia, K.V., J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.
- <sup>3</sup>SAS Institute Inc. *SAS/STAT® 9.2. Users' Guide*. North Carolina, 2008.
- <sup>4</sup>Hansen, M., A. Mukherjee, D. Knorr, and D. Howell. Effect of T-TMA on Capacity and Delay at Los Angeles International Airport. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1788, 2002, pp. 43-48.
- <sup>5</sup>Klein, A., R. Jehlen, and D. Liang. Weather Index with Queuing Component for National Airspace System Performance Assessment. In *Proceedings of 7<sup>th</sup> USA-Europe ATM R&D Seminar*, Barcelona, Spain, 2007.
- <sup>6</sup>Duda, R.O., P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley and Sons, Inc., New York, 2001.
- <sup>7</sup>Hoffman, R., J. Krozel, S. Penny, A. Roy, and K. Roth. A Cluster Analysis to Classify Days in the National Airspace System. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Austin, Texas, 2003.
- <sup>8</sup>Asencio, M. A Clustering Approach for Analysis of Convective Weather Impacting the NAS. Presented at the 12<sup>th</sup> Integrated Communications, Navigation, and Surveillance Conference, Herndon, Virginia, 2012.
- <sup>9</sup>Liu, P., M. Hansen, and A. Mukherjee. Scenario-Based Management of Air Traffic Flow: Developing and Using Capacity Scenario Trees. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1951, 2006, pp. 113-121.
- <sup>10</sup>Enriquez, M. and C. Kurcz. A Simple and Robust Flow Detection Algorithm Based on Spectral Clustering. Presented at the International Conference on Research in Air Transportation. Berkeley, California, 2012.
- <sup>11</sup>Bilimoria, K.D., B. Sridhar, G. Chatterji, and S. Grabbe. Future ATM Concepts Evaluation Tool, In *Air Traffic Control Quarterly*, 9(1), 2001, pp. 1-20.
- <sup>12</sup>Federal Aviation Administration. *The National Playbook*. URL: <http://www.fly.faa.gov/PLAYBOOK/pbindex.html>. Accessed on June 20, 2012.
- <sup>13</sup>U.S. Census Bureau. *Census Regions and Divisions of the United States*. U.S. Department of Commerce. URL: [www.census.gov/geo/www/us\\_regdiv.pdf](http://www.census.gov/geo/www/us_regdiv.pdf). Accessed on June 20, 2012.
- <sup>14</sup>Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. In *Journal of American Statistical Association*, 58(301), 1963, pp. 236-244.